

Fixed Rank Kriging for Cellular Coverage Analysis

Hajer Braham, Sana Ben Jemaa, Gersende Fort, Eric Moulines and Berna Sayrac

Abstract—Coverage planning and optimization is one of the most crucial tasks for a radio network operator. Efficient coverage optimization requires accurate coverage estimation. This estimation relies on geo-located field measurements which are gathered today during highly expensive drive tests (DT); and will be reported in the near future by users' mobile devices thanks to the 3GPP Minimizing Drive Tests (MDT) feature [1]. This feature consists in an automatic reporting of the radio measurements associated with the geographic location of the user's mobile device. Such a solution is still costly in terms of battery consumption and signaling overhead. Therefore, predicting the coverage on a location where no measurements are available remains a key and challenging task. This paper describes a powerful tool that gives an accurate coverage prediction on the whole area of interest: it builds a coverage map by spatially interpolating geo-located measurements using the Kriging technique. The paper focuses on the reduction of the computational complexity of the Kriging algorithm by applying Fixed Rank Kriging (FRK). The performance evaluation of the FRK algorithm both on simulated measurements and real field measurements shows a good trade-off between prediction efficiency and computational complexity. In order to go a step further towards the operational application of the proposed algorithm, a multicellular use-case is studied. Simulation results show a good performance in terms of coverage prediction and detection of the best serving cell.

Keywords—Wireless Network, Coverage Map, Radio Environment Map, Spatial Statistics, Fixed Rank Kriging, Expectation-Maximization algorithm.

I. INTRODUCTION

Coverage planning and optimization is one of the most crucial tasks for a radio network operator. Efficient coverage optimization requires accurate coverage estimation. This estimation relies on geo-located field measurements, gathered today during highly expensive drive tests (DT) and will be reported in the near future by users' mobile devices thanks to the 3GPP Minimization of Drive Tests (MDT) feature standardized since Release 9 [2]. The radio measurements together with the best possible geo-location will be then automatically reported to the network by the user's mobile device. Thanks to the integration of Global Positioning System (GPS) in the new generation of users' mobile devices, the geo-location information is quite accurate. Hence, with MDT, the network operator will soon have at his disposal a rich source of information that provides a greater insight into the end-user perceived quality of service and a better knowledge of the radio environment.

The collection and exploitation of location aware radio measurements was introduced much earlier in the literature in the context of the cognitive radio paradigm [3]. The radio Environmental Map (REM) concept was introduced by Zhao [4] as a database that stores geo-located radio environmental information mainly for opportunistic spectrum access. The REM concept was then extended to an entity that not only stores geo-located radio information but also post processes this information in order to build a complete map. The missing information, namely the considered radio metric in locations where no measurements are available, is then predicted by interpolating the geo-located measurements [5]–[7].

The REM was then studied in the framework of European Telecommunications Standards Institute (ETSI) as a tool for the exploitation of geo-located radio measurements for the radio resource management of mobile wireless networks. A technical report dedicated to the definition of use-cases for building and exploiting the REM gives the following definition [8]: “*The Radio Environment Map (REM) defines a set of network entities and associated protocols that trigger, perform, store and process geo-located radio measurements (received signal strength, interference levels, Quality of Service (QoS) measurements [...]) and network performance indicators. Such measurements are typically performed by user equipments, network entities or dedicated sensors.*” In this ETSI report, several use-cases for REM exploitation in radio resource management are described such as coverage and capacity optimization, and interference management especially for the introduction of a new technology.

Inspired by the geo-statistics area, Kriging technique was applied to REM construction, mainly for coverage prediction and analysis in radio mobile networks [9]–[11]. Bayesian Kriging was first applied to 3G Received Signal Code Power (RSCP) coverage prediction in [9], then to Long Term Evolution (LTE) Reference Signal Received Power (RSRP) coverage analysis in [10]. The description of the bayesian Kriging methodology and the algorithm used in [9], [10], is detailed in [11]. These papers give promising results in terms of performance. However the computational complexity of the algorithm increases cubically with the number of measurement points ($\sim O(N^3)$, where N is the number of measurement points).

In this paper, we aim at providing a method for predicting LTE RSRP coverage map based on MDT data. Given the huge number of measurements that will be reported by mobile terminals with MDT in the near future, reducing the computational complexity of the REM construction becomes crucial. In [12], [13], we used the Fixed Rank Kriging (FRK) introduced by Cressie in [14] (also called in the literature Spatial Random Effects model), as a method to reduce the computational complexity of the Kriging technique applied to radio coverage prediction; the method was evaluated on

H. Braham is with Orange Labs research center, Issy-Les-Moulineaux, France and Télécom ParisTech, Paris, France.

S. Ben Jemaa and B. Sayrac are with Orange Labs research center, Issy-Les-Moulineaux, France.

G. Fort and E. Moulines are with LTCI Télécom ParisTech & CNRS, Paris, France.

simulated data (see [12]) and on real field data (see [13]), both in the situation of a single cell with an omni-directional antenna. In this paper, we go a step further towards operational application of the REM prediction algorithm by considering a multicellular use-case: the directivity of the antennas is introduced in the model, and both the coverage prediction and the good detection of the best serving cell are part of the statistical analysis.

The contribution of this paper can be summarized in the following:

- We describe the FRK algorithm and its adaptation to radio coverage data. It requires an estimation step of the unknown parameters of the model: we show that the method of moments proposed in [14] can not apply and we derive a Maximum Likelihood alternative.
- We extend our model to a multicellular use-case with directive antennas.
- We evaluate the performances of the proposed algorithms both on simulated and real data.

The paper is organized as follows: Section II starts with an overview of the propagation models existing in the literature. Then the statistical parametric model is introduced. The last part is devoted to the parameter estimation: the applicability of the original method is discussed, and an alternative is given. In Section III, the extension to the multicellular use-case is detailed. Then the numerical analysis in the single cell and multicellular use-cases are provided in Section IV. Finally, Section V summarizes the main conclusions.

II. RADIO ENVIRONMENT MAP PREDICTION MODELS

In this section, we give an overview of basic propagation models and give some notations that will be used in the remainder of this paper. Then we introduce a new model for REM construction, which is adapted from the FRK model proposed in [14].

A. Introduction to propagation modeling and notations

A radio propagation model describes a relation between the signal strength, and the locations of the transmitter and the receiver. There are in the literature two different approaches for this description which are respectively derived using analytical and empirical methods [15]. The analytical approach is based on fundamental principals of the radio propagation concept. The empirical one introduces a statistical model and uses a set of observations to fit this model. The advantage of the second approach is the use of actual field measurements to estimate the parameters of the model.

Denote by $Z(x)$ the received power at the receiver end located at $x \in \mathbb{R}^2$, expressed in dB. The *path-loss model*, also called in the literature *the log-distance model*, is among the analytical approaches. It describes $Z(x)$ as a logarithmically decreasing function of the distance $\text{dist}(x)$ between the transmitter location and the receiver location x (see e.g. [15]):

$$Z(x) = p_t - 10\kappa \ln_{10}(\text{dist}(x)), \quad x \in \mathbb{R}^2; \quad (1)$$

p_t is the transmitted power in dB and κ is the path loss exponent. When using this formula to predict the REM,

p_t is considered as known since it is one of the antenna characteristic, and κ depends on the propagation environment. For example, κ is in the order of 2 in free space propagation and it is larger when considering an environment with obstacles (see e.g. [15], [16]).

The model in Eq. (1) does not take into account the fact that two mobile Equipment (ME) equally distant from the base station (BS), may have different environment characteristics. To tackle this bottleneck, empirical approaches based on a statistical modeling of the shadowing effect have been introduced. The *log-normal shadowing model* consists in setting (see [17])

$$Z(x) = p_t - 10\kappa \ln_{10}(\text{dist}(x)) + \sigma_\nu \tilde{\nu}(x), \quad x \in \mathbb{R}^2, \quad (2)$$

where $(\tilde{\nu}(x))_x$, introduced to capture the shadowing effect, is a standard Gaussian variable (note that the terminology “log-normal” comes from the fact that the shadowing term expressed in dB is normally distributed), and $\sigma_\nu > 0$. With this model, the REM prediction at location x is $\hat{Z}(x) = p_t - 10\kappa \ln_{10}(\text{dist}(x))$. The unknown parameters p_t and κ are estimated from measured data, usually by the maximum likelihood estimator (which is also the least-square estimator in this Gaussian case).

Both the models (1) and (2) are large-scale propagation models: they do not consider the small fluctuations of the received power due to the local environment. The *correlated shadowing model* captures these small-scale variations:

$$Z(x) = p_t - 10\kappa \ln_{10}(\text{dist}(x)) + \nu(x), \quad x \in \mathbb{R}^2, \quad (3)$$

where $(\nu(x))_x$ is a zero mean Gaussian process with a parametric spatial covariance function $(C(x, x'))_{x, x'}$. This model implies that two signals $Z(x), Z(x')$ at different locations x, x' are correlated, with covariance equal to $C(x, x')$. The REM prediction formula based on the model (3) is known in the literature as the Kriging (see e.g. [18]): the prediction $\hat{Z}(x)$ is the conditional expectation of $Z(x)$ given the measurements. It depends linearly on these measurements (see [18, Eq. (3.2.12)]) and involves a computational cost $O(N^3)$, where N is the number of measurement points. Here again, the prediction necessitates the estimation of the parameters: different parameter estimation approaches were proposed (see e.g. [18], [19] for maximum likelihood, or [11], [18] for a Bayesian approach). This model was applied to REM interpolation in [11], [19], [20] and this technique has proved to realize accurate prediction performances.

All the models above assume that the antennas are omni-directional. Nevertheless, in macro-cellular networks, operators usually deploy directional antennas. Hence, the received power depends also on the direction of reception. To fit the model to this new constraint, several papers proposed to modify the model (2) by adding a term $\bar{G}(x)$ depending on the mobile location x and modeling the antenna gain (see e.g. [21], [22]): for $x \in \mathbb{R}^2$,

$$Z(x) = p_t - 10\kappa \ln_{10}(\text{dist}(x)) + \bar{G}(x) + \nu(x). \quad (4)$$

Different gain functions \bar{G} are proposed, depending on the antenna used for the transmission (for example, a polar antenna, a sectorial antenna, ...); see e.g. [22]–[24]. The function \bar{G}

depends on parameters which are usually considered known; we will allow the function \bar{G} to depend on unknown parameters to be calibrated from the observations. In this paper, we will extend the model (4) by considering a correlated spatial noise $\nu(x)$.

B. Fixed Rank Kriging prediction model

For $x \in \mathbb{R}^2$, $Z(x)$ is assumed of the form

$$Z(x) = p_t - 10\kappa \ln_{10} \text{dist}(x) + \varsigma G(x) + \mathbf{s}(x)^T \boldsymbol{\eta}, \quad (5)$$

where $\mathbf{s} : \mathbb{R}^2 \rightarrow \mathbb{R}^r$ collects r deterministic spatial basis functions and $\boldsymbol{\eta}$ is a \mathbb{R}^r -valued zero mean Gaussian vector with covariance matrix \mathbf{K} . A^T denotes the transpose of the matrix A and by convention, the vectors are column-vectors. $p_t - 10\kappa \ln_{10} \text{dist}(x) + \varsigma G(x)$ describes the large scale spatial variation (i.e. the trend) and the random process $(\mathbf{s}(x)^T \boldsymbol{\eta})_x$ is a smooth small-scale spatial variation. In practice, the number of basis functions r and the basis functions \mathbf{s} are chosen by the user (see [14, Section 4] and Section IV-B1 below). It is assumed that the function G is known: in the case of an omnidirectional antenna, G is the null function, and for directional antenna we give an example in Section III.

We have N measurement points y_1, \dots, y_N modeled as the realization of the observation vector $\mathbf{Y} = (Y(x_1), \dots, Y(x_N))^T$ at known locations x_1, \dots, x_N and defined as follows

$$Y(x) = Z(x) + \sigma \varepsilon(x), \quad x \in \mathbb{R}^2. \quad (6)$$

$(\varepsilon(x))_x$ is assumed to be a zero mean standard Gaussian process, it incorporates the uncertainties of the measurement technique. $\boldsymbol{\eta}$ and $(\varepsilon(x))_x$ are assumed to be independent so that the covariance matrix of \mathbf{Y} is given by

$$\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_N + \mathbf{S} \mathbf{K} \mathbf{S}^T, \quad (7)$$

where $\mathbf{S} = (\mathbf{s}(x_1), \dots, \mathbf{s}(x_N))^T$ is the $N \times r$ matrix, and \mathbf{I}_N denotes the $N \times N$ identity matrix. This model implies that the conditional distribution of $(Z(x))_x$ given the observations \mathbf{Y} is a Gaussian process. Its expectation and covariance functions are respectively given by (see e.g. [25, Appendix A.2])

$$x \mapsto t^T(x) \boldsymbol{\alpha} + \mathbf{s}(x)^T \mathbf{K} \mathbf{S}^T \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \mathbf{T} \boldsymbol{\alpha}), \quad (8)$$

$$(x, x') \mapsto \mathbf{s}^T(x) \mathbf{K} \mathbf{s}(x') - \mathbf{s}(x)^T \mathbf{K} \mathbf{S}^T \boldsymbol{\Sigma}^{-1} \mathbf{S} \mathbf{K} \mathbf{s}(x'), \quad (9)$$

$$\text{where } \mathbf{T} = \begin{bmatrix} 1 & -10 \ln_{10} \text{dist}(x_1) & G(x_1) \\ \vdots & \vdots & \vdots \\ 1 & -10 \ln_{10} \text{dist}(x_N) & G(x_N) \end{bmatrix},$$

$$\boldsymbol{\alpha} = \begin{bmatrix} p_t \\ \kappa \\ \varsigma \end{bmatrix}, \quad t(x) = \begin{bmatrix} 1 \\ -10 \ln_{10} \text{dist}(x) \\ G(x) \end{bmatrix}.$$

We use the mean value (8) as the estimator $\hat{Z}(x)$ for the unknown quantity $Z(x)$. Note that the estimation of $(Z(x_1), \dots, Z(x_N))^T$ is not \mathbf{Y} since at locations where we have measurements, the prediction technique (8) acts as a denoising algorithm. The prediction formula (8) involves the

inversion of the matrix $\boldsymbol{\Sigma}$. By using standard matrix formulas (see e.g. [26, Section 1.5, Eq. (18)]) we have

$$\boldsymbol{\Sigma}^{-1} = \sigma^{-2} \mathbf{I}_N - \sigma^{-2} \mathbf{S} \left\{ \sigma^2 \mathbf{K}^{-1} + \mathbf{S}^T \mathbf{S} \right\}^{-1} \mathbf{S}^T. \quad (10)$$

The key property of this FRK model is that it only requires the inversion of $r \times r$ matrices. Therefore, the computational cost for the REM prediction is $O(r^2 N)$ which is a drastic reduction when compared to the classical Kriging in situations when N is large. The prediction formula also requires the knowledge of $(\boldsymbol{\alpha}, \sigma^2, \mathbf{K})$. The goal of the following section is to address the estimation of these parameters.

C. Parameter estimation of the Fixed Rank Kriging model

We first expose the method described in the original paper devoted to the FRK model [14]. We also provide a rigorous proof of some weaknesses of this estimation technique pointed out in [27] through numerical experiments. We then propose a second method which is more robust.

1) *Parameter estimation by a method of moments:* In [14], $\boldsymbol{\alpha}$ is estimated by the weighted least squares estimator: given an estimation $(\hat{\sigma}^2, \hat{\mathbf{K}})$ of (σ^2, \mathbf{K}) which yields an estimation $\hat{\boldsymbol{\Sigma}}$ of $\boldsymbol{\Sigma}$ (see Eq. (7)), we have $\hat{\boldsymbol{\alpha}}_{\text{WLS}} = (\mathbf{T}^T \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{T})^{-1} \mathbf{T}^T \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{Y}$. Parameters σ^2 and \mathbf{K} are estimated by a method of moments: the N observations are replaced with M “pseudo-observations” located at x'_1, \dots, x'_M in \mathbb{R}^2 . For each $i = 1, \dots, M$, a pseudo-observation is constructed as the average of the initial observations $Y(x_\ell)$, $\ell = 1, \dots, N$ which are in a neighborhood of x'_i . The parameter M is chosen by the user such that $r < M \ll N$. An empirical $M \times M$ covariance matrix $\hat{\boldsymbol{\Sigma}}_M$ is then associated to these pseudo-observations; it is easily invertible due to its reduced dimensions. Finally, the same “binning” technique is applied to the matrix \mathbf{S} which yields a $M \times r$ matrix \mathbf{S}_M (see [14, Section 3.3.] for a detailed construction of $\hat{\boldsymbol{\Sigma}}_M$ and \mathbf{S}_M ; see also Appendix A below for a partial description). σ^2, \mathbf{K} are then estimated by (see [14, Eq. (3.10)] applied with $\hat{\mathbf{V}} = \mathbf{I}_M$ and $\hat{\mathbf{S}} = \mathbf{S}_M$)

$$\hat{\sigma}^2 = \frac{\text{Tr} \left((\mathbf{I}_M - \mathbf{Q} \mathbf{Q}^T) \hat{\boldsymbol{\Sigma}}_M \right)}{\text{Tr} (\mathbf{I}_M - \mathbf{Q} \mathbf{Q}^T)}, \quad (11)$$

$$\hat{\mathbf{K}} = \mathbf{R}^{-1} \mathbf{Q}^T (\hat{\boldsymbol{\Sigma}}_M - \hat{\sigma}^2 \mathbf{I}_M) \mathbf{Q} (\mathbf{R}^{-1})^T, \quad (12)$$

where Tr denotes the trace and $\mathbf{S}_M = \mathbf{Q} \mathbf{R}$ is the orthogonal-triangular decomposition of \mathbf{S}_M (\mathbf{Q} is a $M \times r$ matrix which contains the first r columns of a unitary matrix and \mathbf{R} is an invertible upper triangular matrix). These estimators are obtained by fitting $\sigma^2 \mathbf{I}_M + \mathbf{S}_M \mathbf{K} \mathbf{S}_M^T$ to $\hat{\boldsymbol{\Sigma}}_M$, solving the optimization problem $\min_{\sigma^2, \mathbf{K}} \|\hat{\boldsymbol{\Sigma}}_M - \sigma^2 \mathbf{I}_M - \mathbf{S}_M \mathbf{K} \mathbf{S}_M^T\|$ where in this equation, $\|\cdot\|$ denotes the Froebenius norm (to have a better intuition of this strategy, compare this criterion to Eq. (7)). $\hat{\mathbf{K}}$ has to be positive definite since it estimates an invertible covariance matrix. In [27], the authors observe through numerical examples that the estimator (12) is a singular covariance matrix (hence, they introduce an “eigenvalue lifting” procedure to modify (12) and obtain a positive

definite matrix (see [27, Section 3.2.]). We identify sufficient conditions for this empirical observation to be always valid. More precisely, we establish in Appendix A the following,

Proposition 1: Assume that \mathbf{S}_M is a full rank matrix and let $\mathbf{S}_M = \mathbf{Q}\mathbf{R}$ be its orthogonal-triangular decomposition (\mathbf{Q} is a $M \times r$ matrix which collects the first r columns of a unitary matrix). Denote by $(\lambda_j)_j$ the eigenvalues of $\hat{\Sigma}_M$ and \mathbf{V}_j the eigenspace of λ_j . Then

- (i) $\hat{\Sigma}_M$ is positive semi-definite.
- (ii) $\hat{\sigma}^2$ given by (11) is lower bounded by $\inf_{j: \exists v \in \mathbf{V}_j, \|\mathbf{Q}^T v\| < \|v\|} \lambda_j$.
- (iii) $\hat{\mathbf{K}}$ given by (12) is positive definite iff $\hat{\sigma}^2 \in [0, \lambda_{\min}(\mathbf{Q}^T \hat{\Sigma}_M \mathbf{Q})]$ where $\lambda_{\min}(A)$ denotes the minimal eigenvalue of A .

We also give in Appendix A a sufficient condition which implies that the minimal eigenvalue (say λ_1) of $\hat{\Sigma}_M$ is positive. If there exists $v \in \mathbf{V}_1$ such that $\|\mathbf{Q}^T v\| = \|v\|$ then $\mathbf{Q}^T v$ is an eigenvector of $\mathbf{Q}^T \hat{\Sigma}_M \mathbf{Q}$ associated to the eigenvalue λ_1 (observe indeed that if $\|\mathbf{Q}^T v\| = \|v\|$, then there exists $\mu \in \mathbb{R}^r$ such that $v = \mathbf{Q}\mu$ and this vector satisfies $\mu = \mathbf{Q}^T v$). Therefore, if $\lambda_1 > 0$ and for any $v \in \mathbf{V}_1$, $\|\mathbf{Q}^T v\| = \|v\|$ then Proposition 1 implies that $\hat{\mathbf{K}}$ given by (12) can not be positive definite.

2) *Parameter estimation by Maximum Likelihood:* We propose to estimate the parameters by the Maximum Likelihood Estimator (MLE), following an idea close to that of [28], [29]. Observe from (5) and (6) that $\mathbf{Y} = \mathbf{T}\alpha + \mathbf{S}\eta + \sigma\epsilon$ with $\epsilon = (\epsilon(x_1), \dots, \epsilon(x_N))^T$. This equation shows that from \mathbf{Y} , it is not possible to estimate a general covariance matrix \mathbf{K} since roughly speaking, \mathbf{Y} is obtained from a single realization of a Gaussian vector η with covariance matrix \mathbf{K} . Therefore, we introduce a parametric model for this covariance matrix, depending on some vector v of low dimension: we will write $\mathbf{K}(v)$. We give an example of such a parametric family in Section IV-B2; see also [25, Chapter 4].

Since η and $(\epsilon(x))_x$ are independent processes, \mathbf{Y} is a \mathbb{R}^N -valued Gaussian vector with mean $\mathbf{T}\alpha$ and with covariance matrix $\Sigma = \sigma^2 \mathbf{I}_N + \mathbf{S}\mathbf{K}(v)\mathbf{S}^T$. Therefore the log-likelihood $L_{\mathbf{Y}}(\theta)$ of the observations \mathbf{Y} given the parameters $\theta = (\alpha, \sigma^2, v)$ is, up to an additive constant,

$$L_{\mathbf{Y}}(\theta) = -\frac{1}{2} \ln \det(\sigma^2 \mathbf{I}_N + \mathbf{S}\mathbf{K}(v)\mathbf{S}^T) - \frac{(\mathbf{Y} - \mathbf{T}\alpha)^T}{2\sigma^2} \left(\mathbf{I}_N - \mathbf{S} \left\{ \sigma^2 \mathbf{K}^{-1}(v) + \mathbf{S}^T \mathbf{S} \right\}^{-1} \mathbf{S}^T \right) \dots \times (\mathbf{Y} - \mathbf{T}\alpha), \quad (13)$$

where we used (10) for the expression of Σ^{-1} . Maximizing directly the log-likelihood function $\theta \mapsto L_{\mathbf{Y}}(\theta)$ is not straightforward and cannot be computed analytically. We therefore propose a numerical solution based on the Expectation Maximization (EM) algorithm [30]. EM allows the computation of the MLE in latent data models; in our framework, the latent variable is η . It is an iterative algorithm which produces a sequence $(\theta_{(l)})_{l \geq 0}$ satisfying $L_{\mathbf{Y}}(\theta_{(l+1)}) \geq L_{\mathbf{Y}}(\theta_{(l)})$. This property is fundamental for the proof of convergence of any EM sequence [31]. Each iteration of EM consists in two steps:

an Expectation step (E-step) and a Maximization step (M-step). Given the current value $\theta_{(l)}$ of the parameter, the E-step consists in the computation of the expectation of the log-likelihood of (\mathbf{Y}, η) under the conditional distribution of η given \mathbf{Y} for the current value of the parameter $\theta_{(l)}$:

$$Q(\theta; \theta_{(l)}) = \mathbb{E} [\ln \Pr(\mathbf{Y}, \eta; \theta) | \mathbf{Y}; \theta_{(l)}],$$

where $\theta \mapsto \Pr(\mathbf{Y}, \eta; \theta)$ is the likelihood of (\mathbf{Y}, η) . In the M-step, the parameter is updated as the value maximizing $\theta \mapsto Q(\theta; \theta_{(l)})$ or as any value $\theta_{(l+1)}$ satisfying

$$Q(\theta_{(l+1)}; \theta_{(l)}) \geq Q(\theta_{(l)}; \theta_{(l)}). \quad (14)$$

The E- and M-steps are repeated until convergence, which in practice may mean when the difference between $\|\theta_{(l)} - \theta_{(l+1)}\|$ changes by an arbitrarily small amount determined by the user (see e.g. [30, Chapter 3]). In our framework, we have

$$Q(\theta; \tilde{\theta}) = -\frac{N}{2} \ln(\sigma^2) - \frac{1}{2} \ln(\det(\mathbf{K}(v))) - \frac{1}{2\sigma^2} \|\mathbf{Y} - \mathbf{T}\alpha\|^2 - \frac{1}{2} \text{Tr} \left(\left(\frac{\mathbf{S}^T \mathbf{S}}{\sigma^2} + \mathbf{K}^{-1}(v) \right) \mathbb{E} [\eta \eta^T | \mathbf{Y}; \tilde{\theta}] \right) + \frac{1}{\sigma^2} (\mathbf{Y} - \mathbf{T}\alpha)^T \mathbf{S} \mathbb{E} [\eta | \mathbf{Y}; \tilde{\theta}], \quad (15)$$

where (see e.g. [12, Appendix C])

$$\mathbb{E} [\eta | \mathbf{Y}; \tilde{\theta}] = \left(\mathbf{S}^T \mathbf{S} + \tilde{\sigma}^2 \mathbf{K}^{-1}(\tilde{v}) \right)^{-1} \mathbf{S}^T (\mathbf{Y} - \mathbf{T}\tilde{\alpha}),$$

$$\text{cov} [\eta | \mathbf{Y}; \tilde{\theta}] = \left(\frac{\mathbf{S}^T \mathbf{S}}{\tilde{\sigma}^2} + \mathbf{K}^{-1}(\tilde{v}) \right)^{-1}.$$

The update formulas of the parameters (α, σ^2) are given by (see e.g. [12, Appendix B] for the proof)

$$\alpha_{(l+1)} = \left(\mathbf{T}^T \mathbf{T} \right)^{-1} \mathbf{T}^T (\mathbf{Y} - \mathbf{S} \mathbb{E} [\eta | \mathbf{Y}; \theta_{(l)}]),$$

$$\sigma_{(l+1)}^2 = \frac{1}{N} \mathbb{E} [\|\mathbf{Y} - \mathbf{T}\alpha_{(l+1)} - \mathbf{S}\eta\|^2 | \mathbf{Y}; \theta_{(l)}].$$

With this choice, we have $Q(\alpha_{(l+1)}, \sigma_{(l+1)}^2, v; \theta_{(l)}) \geq Q(\theta_{(l)}; \theta_{(l)})$, for any v . The update of v is specific to each parametric model for \mathbf{K} . Upon noting that the first order derivative of $v = (v_1, \dots, v_p) \mapsto Q(\alpha, \sigma^2, v; \theta_{(l)})$ w.r.t. v_k is given by

$$-\frac{1}{2} \text{Tr} \left(\mathbf{K}^{-1}(v) \frac{\partial \mathbf{K}(v)}{\partial v_k} \right) + \frac{1}{2} \text{Tr} \left(\mathbf{K}^{-1}(v) \mathbb{E} [\eta \eta^T | \mathbf{Y}; \theta_{(l)}] \mathbf{K}^{-1}(v) \frac{\partial \mathbf{K}(v)}{\partial v_k} \right), \quad (16)$$

$v_{(l+1)}$ can be defined as the unique root of this gradient whenever it is the global maximum. Another strategy is to perform one iteration of a Newton-Raphson algorithm starting from $v_{(l)}$ with a step size chosen in order to satisfy the EM condition (14). See e.g. [30, Section 4.14] for EM combined with Newton-Raphson procedures. In Section IV-B2, we will give an example of structured covariance matrix and will derive the Newton-Raphson strategy to update one of the parameters.

III. REM EXTENDED TO MULTICELLULAR NETWORK

We now consider a multicellular LTE network. In real network, UEs measure the received power of several BSs in order to choose the best serving one: the UE, this procedure is called the *cell selection*. In LTE, cell selection is applied by comparing the instant measured RSRP from all potential cells and choosing the cell providing the highest RSRP value [32]. In this section, we adapt the FRK model and the REM prediction technique described in Section II-B in order to address this multicellular use-case.

We assume that the reported measurements correspond to the RSRP of the best serving cell: each measurement consists in the RSRP measure, the location information and the corresponding cell identifier (CID). The received power $Z_i(x)$ from the i -th BS at location x is given by $Z_i(x) = 0$ if $x \notin \mathcal{D}_i$ and if $x \in \mathcal{D}_i$,

$$Z_i(x) = p_{t,i} - 10\kappa_i \ln_{10}(\text{dist}_i(x)) + \varsigma_i G_i(x) + \mathbf{s}_i(x)^T \boldsymbol{\eta}_i \quad (17)$$

where $\mathcal{D}_i \subseteq \mathbb{R}^2$, $p_{t,i}$ is the transmitted power of the i -th BS, κ_i is the path loss exponent corresponding to the i -th BS and $\text{dist}_i(x)$ is the distance from x to the i -th BS. We can choose $\mathcal{D}_i \neq \mathbb{R}^2$ to model geographic area which are not covered by the i -th BS. $\boldsymbol{\eta}_i$ is a Gaussian variable with zero mean and covariance matrix \mathbf{K}_i . $\mathbf{s}_i(x) : \mathbb{R}^2 \rightarrow \mathbb{R}^{r_i}$ collects r_i deterministic spatial basis functions.

$\varsigma_i G_i(x)$ is the antenna gain which depends on the mobile location x . In our use-case, the antennas used for each BS are tri-sectorial; we use a typical antenna pattern proposed in the 3GPP standard [1] with a horizontal gain only since we are using a 2-dimensional model:

$$G_i(x) = -\min \left[12 \left(\frac{\psi_{x,i}}{\psi_{3dB}} \right)^2, A_m \right], \quad (18)$$

where $\psi_{x,i}$ is the angle between the UE location x , and the i -th BS antenna azimuth. ψ_{3dB} denotes the angle at which the antenna efficiency is 50% and A_m is the maximum antenna gain. For a tri-sectorial antenna, the parameter ψ_{3dB} is usually taken equal to 65° and $A_m = 30\text{dB}$.

We have N_i observations $Y_i(x)$ having the i -th BS as the best serving cell. They are located at $x_{1,i}, \dots, x_{N_i,i}$ and are noisy measurements of $Z_i(x)$: $Y_i(x) = Z_i(x) + \sigma_i \varepsilon_i(x)$ where $(\varepsilon_i(x))_x$ is a zero mean standard Gaussian process, independent of $\boldsymbol{\eta}_i$. Following the same lines as in section II-B, we define the $N_i \times 1$ column vector $\mathbf{Y}_i = (Y_i(x_{1,i}), \dots, Y_i(x_{N_i,i}))^T$, and have $\mathbf{Y}_i = \mathbf{T}_i \boldsymbol{\alpha}_i + \mathbf{S}_i \boldsymbol{\eta}_i + \sigma_i \boldsymbol{\varepsilon}_i$ where

$$\mathbf{T}_i = \begin{bmatrix} 1 & -10 \ln_{10}(\text{dist}_i(x_{1,i})) & G_i(x_{1,i}) \\ \vdots & \vdots & \vdots \\ 1 & -10 \ln_{10}(\text{dist}_i(x_{N_i,i})) & G_i(x_{N_i,i}) \end{bmatrix},$$

$$\boldsymbol{\alpha}_i = \begin{bmatrix} p_{t,i} \\ \kappa_i \\ \varsigma_i \end{bmatrix}, \quad \boldsymbol{\varepsilon}_i = \begin{bmatrix} \varepsilon_i(x_{1,i}) \\ \vdots \\ \varepsilon_i(x_{N_i,i}) \end{bmatrix}.$$

The parameters $p_{t,i}$, κ_i , σ_i , ς_i and \mathbf{K}_i are unknown and are

estimated from \mathbf{Y}_i by applying the EM technique described in Section II-C (see also Section IV-B for the implementation).

For any x such that $x \in \mathcal{D}_i$, set $\hat{Z}_i(x) = \mathbb{E}[Z_i(x)|\mathbf{Y}_i]$, the expression of which can easily be adapted from (8). In the multicellular case, the inter-site shadowing correlation can be explained by a partial overlap of the large-scale propagation medium as explained in [33]. Hence, for any x such that $x \in \mathcal{D}_i$, we write $Z_i(x) = Z'_i(x) + W(x)$, where $W(x)$ is the random cross-correlated shadowing term which depends only on the mobile location (also called overlapping propagation term) and $Z'_i(x)$ is the random correlated shadowing related to the i -th BS at the location x (also called non-overlapping propagation term). As explained in [33], the r.v. $(Z'_i(x))_i$ are independent, which implies that the probability that a UE located at x is attached to the i -th BS (which is denoted by $\text{CID}(x) = i$) is given by

$$\mathbb{P}(\text{CID}(x) = i) = \mathbb{E} \left[\prod_{j \neq i: x \in \mathcal{D}_j} \mathbb{1}_{Z_j(x) \leq Z_i(x)} \right]. \quad (19)$$

A simple approximation consists in approximating this expectation by

$$\prod_{j \neq i: x \in \mathcal{D}_j} \mathbb{1}_{\hat{Z}_j(x) \leq \hat{Z}_i(x)}.$$

This yields the estimation rules for the CID and the RSRP value at x

$$\widehat{\text{CID}}(x) = \underset{j: x \in \mathcal{D}_j}{\text{argmax}} \hat{Z}_j(x),$$

$$\hat{Z}(x) = \hat{Z}_{\widehat{\text{CID}}(x)}(x) = \max_{j: x \in \mathcal{D}_j} \hat{Z}_j(x).$$

IV. APPLICATIONS TO CELLULAR COVERAGE MAP

A. Data sets description

For the single cell use-case, we consider a simulated data set and a real data set. The first data set consists of simulated measurement points generated with a very accurate planning tool, which uses a sophisticated ray-tracing propagation model developed for operational network planning [34]. This data is considered as the ground-truth of the coverage in the area of interest. The collected data set corresponds to the LTE RSRP values in an urban scenario located in the Southwest of Paris (France). The environment is covered by a macro-cell with an omni-directional antenna. These measurement points are located on a 1000 m \times 1000 m surface, regularly spaced on a cartesian grid consisting of 5 m \times 5 m squares; this yields a total of 40401 measurement points (see Fig. 1a, where the antenna location is (595 416 m, 2 425 341 m)). In order to model the noise measurements, a zero mean Gaussian noise with variance equal to 3 dB is added to the simulated measurements. This yields what we called in Section II the process $\{Y(x), x \in \mathcal{D}\}$, where $\mathcal{D} \subset \mathbb{R}^2$.

The second data set corresponds to real measurement points reported from Drive Tests (DT) done by Orange France teams, in a rural area located in southwestern France. The BS is about 30 m height and covers an area of 22 km \times 10 km. 7800 measurement points have been collected in the 800

MHz frequency band using a typical user's mobile device connected to a software tool for data acquisition. The locations of the measurement points are shown on Fig. 1b - note that they are along the roads and the antenna is located at (408 238 m, 1 864 600 m). For the multicellular use-case, we consider a simulated data set provided by the aforementioned Orange planning tool. This planning tool calculates RSRP values in a sub-urban environment shown in Fig. 2a, consisting of 12 antennas grouped into 4 sites of 3 directional antennas. The inter-site distance is bigger than 1 km. The antennas are tri-sector. The RSRP values are computed over a regular grid of size 25 m×25 m over a 12.4 km² geographic area, which results in a total of 20 008 locations; and it is realized over a 2.6 GHz frequency band. The planning tool returns, at each location of the regular grid, both the RSRP value and the ID of the best serving cell. Fig. 2b displays the RSRP values and Fig. 2c shows the best serving cell map where each color corresponds to a cell coverage area.

B. EM implementation

1) *Choice of the basis functions s* : The basis functions $x \mapsto s(x) = (S_1(x), \dots, S_r(x))$ and their number r both control the complexity and the accuracy of the FRK prediction technique. Following the suggestions in [14], we choose the l -th basis function $x \mapsto S_l(x)$ as a symmetric function centered at locations x'_l : S_l is a bi-square function defined as

$$S_l(x) = \begin{cases} \left[1 - (\|x - x'_l\| / \tau)^2\right]^2, & \text{if } \|x - x'_l\| \leq \tau, \\ 0, & \text{otherwise} \end{cases} \quad (20)$$

The parameter τ controls the support of the function. In the numerical applications below, the centers of the basis functions x'_l and their number r are chosen as follows: r_{\max} functions are located on a Cartesian grid where the elements are $\tau \times \tau$ squares covering the whole geographic area of interest. Then, for each function S_l , if none of the N locations x_1, \dots, x_N is in a τ -neighborhood of the center x'_l , this function is removed. The number of the remaining basis function is r . On Fig. 3a and Fig. 3b, we show the locations of the N observations (red circle) and the locations of the r basis function centers (blue crosses) for two different data sets. In Fig. 3a, $\tau = 100$ m and $r = r_{\max}$ (and $N = 2000$) while in Fig. 3b, $\tau = 250$ m, $r_{\max} = 2660$ and $r = 467$.

2) *Structured covariance matrix K* : Several examples of structured covariance matrix K can be chosen. In the radio cellular context, the shadowing term can be modeled as a zero-mean Gaussian random variable with an exponential correlation model [35]. Thus, K is given by

$$K(\beta, \phi) = \frac{\tilde{K}(\phi)}{\beta}, \quad (21)$$

$$\text{with } \tilde{K}_{i,j}(\phi) = \exp\left(-\frac{\|x'_i - x'_j\|}{\exp(\phi)}\right), \quad (22)$$

where $\|x'_i - x'_j\|$ is the Euclidean distance between the two locations x'_i and x'_j (related to the basis functions, see Section IV-B1). $1/\beta$ and $\exp(\phi)$ are respectively the variance of

η_l , $1 \leq l \leq r$; and a rate of decay of the correlation (the choice of the parametrization $\exp(\phi)$ avoids the introduction of a constraint of sign when estimating ϕ). We therefore have $v = (\beta, \phi) \in \mathbb{R}_+^+ \times \mathbb{R}$. For this specific parametric matrix (21-22), a possible update of the parameters (β, ϕ) which ensures the monotonicity property of the EM algorithm is (see e.g. [12, Appendix B]): $\beta_{(l+1)} = r / \text{Tr}(\tilde{K}_{(l)}^{-1} \mathcal{V}_{(l)})$ and

$$\phi_{(l+1)} = \phi_{(l)} - \frac{a_{(l)}}{\mathcal{H}_{(l)}} \dots \times \text{Tr}\left(\left(\beta_{(l+1)} \tilde{K}_{(l)}^{-1} \mathcal{V}_{(l)} - I_r\right) \tilde{K}_{(l)}^{-1} \Delta \circ \tilde{K}_{(l)}\right)$$

where $\tilde{K}_{(l)}$ is a shorthand notation for $\tilde{K}(\phi_{(l)})$, Δ is the $r \times r$ matrix with entries $(\|x'_i - x'_j\|)_{ij}$, $\mathcal{V}_{(l)}$ is a shorthand notation for $\mathbb{E}[\eta\eta^T | \mathbf{Y}; \theta_{(l)}]$, \circ denotes the Hadamard product and

$$\begin{aligned} \mathcal{H}_{(l)} = & -\text{Tr}\left(\tilde{K}_l^{-1} \Delta \circ \tilde{K}\left(\beta_{(l+1)} \tilde{K}_l \mathcal{V}_{(l)} - I_r\right)\right) \\ & + \exp(-\phi_{(l)}) \text{Tr}\left(\tilde{K}_l^{-1} \Delta \circ \Delta \circ \tilde{K}_l\left(\beta_{(l+1)} \tilde{K}_l \mathcal{V}_{(l)} - I_r\right)\right) \\ & + \exp(-\phi_{(l)}) \text{Tr}\left(\left(\tilde{K}_l^{-1} \Delta \circ \tilde{K}_l\right)^2 \left(I_r - 2\beta_{(l+1)} \tilde{K}_l \mathcal{V}_{(l)}\right)\right); \end{aligned}$$

$a_{(l)} \in (0, 1)$ is chosen so that $Q(\theta_{(l+1)}; \theta_{(l)}) \geq Q(\theta_{(l)}; \theta_{(l)})$.

3) *EM convergence*: EM converges whatever the initial value $\theta_{(0)}$ (see [31]); the limiting points of the EM sequences are the stationary points of the log-likelihood of the observations \mathbf{Y} . We did not observe that the initialization $\theta_{(0)}$ plays a role on the limiting value of our EM runs. A natural initial value for α is the Ordinary Least Square estimator given by $\alpha_{(0)} = \left(\mathbf{T}^T \mathbf{T}\right)^{-1} \mathbf{T}^T \mathbf{Y}$. We choose $\phi_{(0)}$ large enough so that the matrix $\tilde{K}(\phi_{(0)})$ looks like the identity matrix; in practice, we choose $\tau / \exp(\phi)$ in the order of 5. Finally, we compute the empirical variance \mathcal{V} of the components of the residual vector $\mathbf{Y} - \mathbf{T}\alpha_{(0)}$ and choose $\beta_{(0)}^{-1} + \sigma_{(0)}^2 = \mathcal{V}$; roughly speaking, we start from a model with uncorrelated shadowing term. The algorithm is stopped when $\|\theta_{(l)} - \theta_{(l-1)}\| < 10^{-5}$ over 100 successive iterations. We report in Table I the values of the parameters at convergence of EM for the simulated data set.

TABLE I. SIMULATED DATA SET, WHEN $\tau = 50$ M, $r = 400$ AND $N = 32000$

$\hat{\sigma}^2$	$\hat{\alpha}$	$1/\hat{\beta}$	$\hat{\phi}$	
18.15	-49.55	2.73	12.5	3.63

C. Prediction Error Analysis for the single cell use-case

Each data set is splitted into a learning set and a test set. Using the data in the learning set, the parameters are estimated by the method described in Section II-C. The performances are then evaluated using the data in the test set. In order to make this analysis more robust to the choice of the learning and test sets, we perform a k -fold cross validation [36] (here, we choose $k = 5$) with a uniform data sampling of the

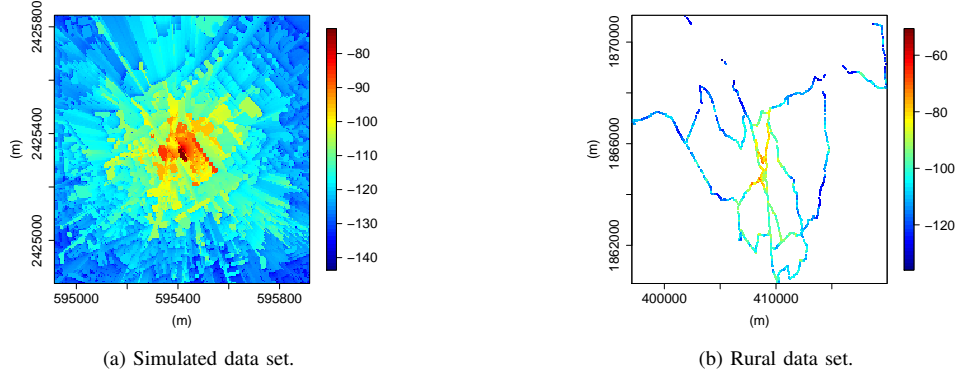


Fig. 1. One cell case: the measurements $(Y(x))_x$.

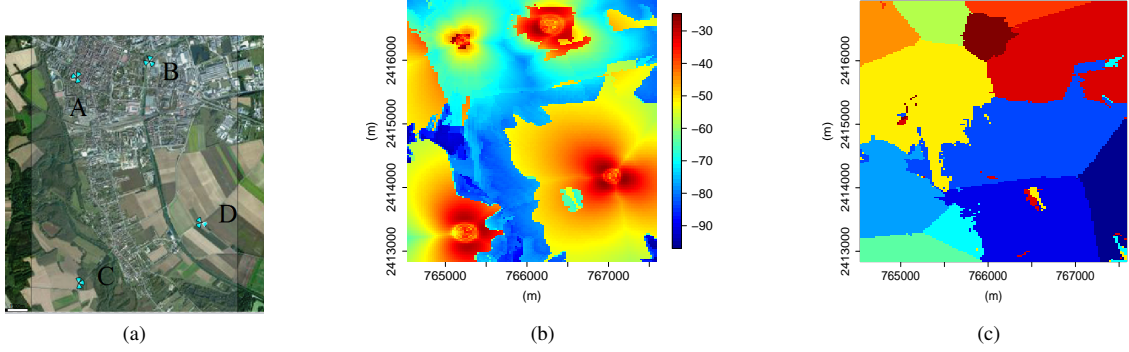


Fig. 2. Multicellular case: (a) BS locations; (b) the simulated RSRP map; (c) measurements grouped in 12 clusters, according to their best serving cell ID

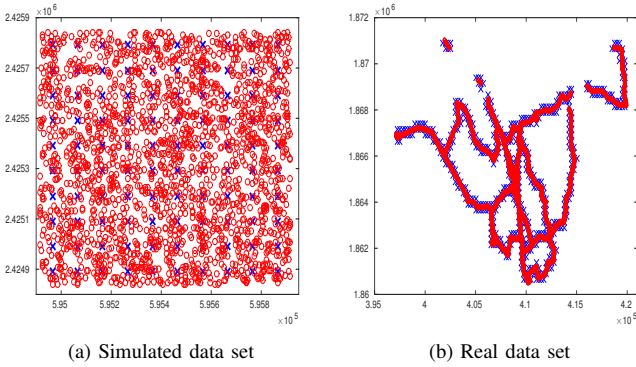


Fig. 3. Locations of the N observations (red circles) and locations of the r centers x'_i (blue crosses) of the basis functions.

subsets (typical values for k are in the range 3 to 10 [25, see Section 5.3.]). Therefore, at each step of this cross-validation procedure, we have a learning set consisting of 80% of the available measurement points (making a learning sets with resp. 32000 and 6000 points for resp. the simulated data set

and the real data set).

In order to evaluate the prediction accuracy, we compare the measurements $Y(x)$ to the predicted values $\hat{Y}(x)$ from the model (6). We consider the locations x in the test set \mathcal{T} . The model (6) implies that the conditional expectation of $Y(x)$ given \mathbf{Y} at such locations x is equal to the conditional expectation of $Z(x)$ given \mathbf{Y} since $\varepsilon(x)$ is independent of \mathbf{Y} . Therefore, for any $x \in \mathcal{T}$, the error (with sign) is $\hat{Y}(x) - Y(x) = \hat{Z}(x) - Y(x)$ where $\hat{Z}(x)$ is given by (8). We evaluate the Root Mean Square Error (RMSE) which is a commonly used prediction error indicator (see e.g. [37]), defined as

$$\text{RMSE} = \left[\frac{1}{|\mathcal{T}|} \sum_{x \in \mathcal{T}} \left(\hat{Y}(x) - Y(x) \right)^2 \right]^{\frac{1}{2}}, \quad (23)$$

where $|\mathcal{T}|$ denotes the number of observations in the test set \mathcal{T} . The RMSE is computed for each of the k successive test sets in the cross-validation analysis. In Tables II and III, we report the mean value of the RMSE over the k partitions and its standard deviation in parenthesis. We compare different strategies for modeling the observations $(Y(x))_x$, for the

parameter estimation of the model and for the prediction:

- Log-Normal: the log-normal shadowing model (see (2)) when the parameters p_t, κ, σ^2 are estimated by MLE. $\hat{Z}(x)$ is given by $\hat{p}_t - 10\hat{\kappa} \log_{10}(\text{dist}(x))$; this method does not depend on r .
- FRK: the FRK model (see section II-B) when the parameters are estimated by MLE (see Sections II-C and IV-B) and $\hat{Z}(x)$ is given by (8), for different values of r .

In tables II and III, we report the mean RMSE over the k splits of the data set and its standard deviation between parenthesis. These tables show that the FRK model improves on the log-

TABLE II. SIMULATED DATA SET: MEAN RMSE AND STANDARD DEVIATION IN PARENTHESIS.

Log-Normal	FRK $r = 1089$	FRK $r = 100$
5.08 (6.08e-02)	3.98 (5.18e-02)	4.67 (4.46e-02)

TABLE III. REAL DATA SET: MEAN RMSE AND STANDARD DEVIATION IN PARENTHESIS

Log-Normal	FRK $r = 1000$	FRK $r = 150$
8.95 (1.46e-01)	3.51 (1.24e-01)	5.57 (6.23e-02)

normal model. For the real data set, it yields a considerably low RMSE (in the order of 3 – 5 dB) when compared to the log-normal shadowing model which has a RMSE in the order of 9 dB. For the simulated data set, we have a similar behavior.

The computational complexity of the FRK approach is essentially related to r , the number of basis functions. On the one hand, the computational cost increases with r and on the other hand, the prediction accuracy increases with r . We report on Fig. 4 the running time and the prediction accuracy measured in terms of mean RMSE over the k splitting of the data set into a learning and a test set, as a function of r ; by convention, the running time is set to 1 when $r = 64$. The plot is obtained with 7 different analysis, obtained with $\tau \in \{30, 40, 50, 60, 80, 100, 120\}$ - or equivalently, $r \in \{1089, 625, 400, 289, 169, 100, 64\}$. It shows that the running time is multiplied by a factor 130 and the prediction accuracy is increased by 20% when moving from $\tau = 120$ ($r = 64$) to $\tau = 30$ ($r = 1089$).

D. Prediction Error Analysis for the multicellular use-case

The data set is splitted into a learning set with 16 000 points and a test set. Based on their best serving cell ID, these 16 000 points are clustered into 12 subsets. The size of these subsets varies between 1000 and 3500. In Fig. 5a a learning subset associated to a given BS is displayed: note that the observations with a given best serving cell ID are not uniformly distributed over the geographical area of interest. We choose the same initial basis functions for the 12 sub-models (defined by Eq.(20) with $\tau = 150$, which yields

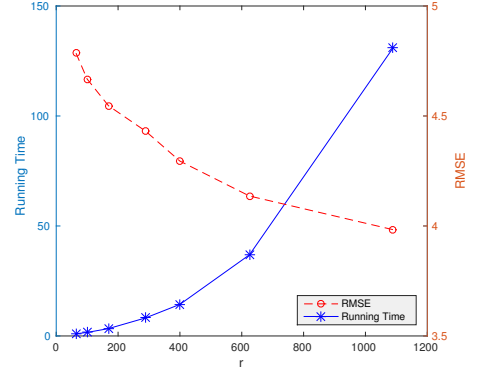


Fig. 4. Simulated Data set: for different values of r , the running time and the mean RMSE

$r_{\max} = 588$). For each sub-model, some of the basis functions are canceled as described in Section IV-B1 (see the blue circles and black dots in Fig. 5a). Fig. 5b shows the path-loss function $x \mapsto \hat{p}_{t,i} - 10\hat{\kappa}_i \ln_{10} \text{dist}_i(x) + \hat{\zeta}_i G_i(x)$: note that, as expected, $T_i \alpha_i$ is bigger in the direction of the antenna spread. In Fig. 5c, we display $\{\hat{Z}_i(x), x \in \mathcal{D}_i\}$. \mathcal{D}_i is defined as the area covering the main direction of the i -th antenna radiation.

The best serving cell ID (CID_{bs}) for any location $x \in \mathcal{D}_i$ is defined as the ID of the BS having the biggest probability that the ME is attached to it at location x as detailed in Eq. 19. Then the predicted received power at location x corresponds to the predicted received power of the best serving cell at that location. For performance evaluation, we first consider an omni-directional antenna model (similar to the one in section IV-C). We compare the predicted cell ID for each location x (that is the index j such that $\hat{Z}(x) = \hat{Z}_j(x)$) to the real one. We obtain an error rate of 53% over the locations x in the test set. When we consider the domain clustering introduced in (17) (the antennas are still assumed to be omnidirectional), the error rate on cell ID selection is 31.23% over the test set locations. Finally, we consider the directional model together with the same domain restriction \mathcal{D}_i . The error rate is drastically decreased to 12.64%. This error rate is expected to further decrease when using real antenna patterns (the impact of approximating real antenna patterns with the 3GPP model is studied for example in [38]).

V. CONCLUSION

In this work we have studied the performance of the FRK algorithm applied to coverage analysis in cellular networks. This method has a good potential when performing prediction using massive data sets (order of thousands and higher) as it offers a good trade-off between prediction quality and computational complexity compared to classical Kriging techniques. This study has been performed using field-like measurements obtained from an accurate planning tool and real field measurements obtained from drive tests. In addition we have adapted the model to a more practical application: we

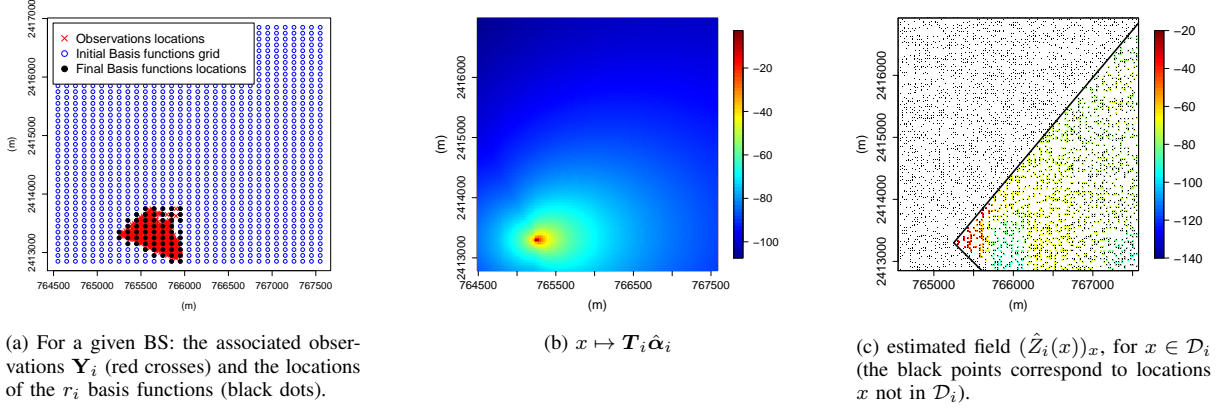


Fig. 5. Multicellular case: results for a given best serving cell ID i

used field-like measurements over several cells with directive antennas. Simulation results show a good performance in terms of coverage prediction and detection of the best serving cell. In future works, we target to further improve this performance by using real antenna patterns. Finally, our ongoing research focuses on extending the model to take into account the location uncertainty and on studying its impact on the prediction performances.

APPENDIX A PROOF OF PROPOSITION 1

We recall some notations introduced in [14, Appendix A], which will be useful for the proof of Proposition 1. For $j = 1, \dots, M$, set $\mathbf{W}_j = (W_{j1}, \dots, W_{jN})^T$, where W_{lj} is the weight associated to the observation $Y(x_j)$ in the neighborhood of the bin center x'_l (see [14] for the expression of these non negative weights). Define the vector of residual $\mathbf{D} = (D_1, \dots, D_N)^T = \mathbf{Y} - \mathbf{T}(\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T \mathbf{Y}$, and associate an aggregated vector of residuals $\bar{\mathbf{D}} = (\bar{D}_1, \dots, \bar{D}_M)^T$ and a weighted square residuals

$$\bar{D}_\ell = \frac{\sum_{i=1}^N W_{\ell i} D_i}{\sum_{i=1}^N W_{\ell i}} = \frac{\mathbf{W}_\ell^T \mathbf{D}}{\mathbf{W}_\ell^T \mathbf{1}_N}, \quad V_\ell = \frac{\sum_{i=1}^N W_{\ell i} D_i^2}{\mathbf{W}_\ell^T \mathbf{1}_N}.$$

$\mathbf{1}_N$ is the $N \times 1$ vector of ones. The $M \times M$ matrix $\hat{\Sigma}_M$ is defined by (see [14, Eq. (A.2)])

$$\hat{\Sigma}_M(l, k) = \bar{D}_\ell \bar{D}_k, \text{ for } l \neq k, \quad \hat{\Sigma}_M(k, k) = V_k. \quad (24)$$

Proof of Proposition 1 (i) Let $\mu = (\mu_1, \dots, \mu_M) \in \mathbb{R}^M$. From (24),

$$\begin{aligned} \mu^T \hat{\Sigma}_M \mu &= \left(\sum_{l=1}^M \mu_l \bar{D}_l \right)^2 + \sum_{l=1}^M \mu_l^2 (V_l - \bar{D}_l^2) \\ &\geq \sum_{l=1}^M \mu_l^2 (V_l - \bar{D}_l^2). \end{aligned}$$

The Jensen's inequality implies that $V_l \geq \bar{D}_l^2$ for any l thus showing that $\mu^T \hat{\Sigma}_M \mu \geq 0$. Note also that this term is positive for any non null vector μ iff $V_l - \bar{D}_l^2 > 0$ for any l . (ii) Since $\hat{\Sigma}_M$ is a covariance matrix, there exists an orthogonal $M \times M$ matrix \mathbf{U} and a diagonal $M \times M$ matrix $\mathbf{\Lambda}$ with diagonal entries $(\lambda_i)_i$ such that $\hat{\Sigma}_M = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T$. Since $\text{Tr}(\mathbf{A}\mathbf{B}) = \text{Tr}(\mathbf{B}\mathbf{A})$, we have

$$\text{Tr}((\mathbf{I}_M - \mathbf{Q}\mathbf{Q}^T) \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T) = \sum_{i=1}^M B_{ii} \lambda_i,$$

where $\mathbf{B} = \mathbf{U}^T (\mathbf{I}_M - \mathbf{Q}\mathbf{Q}^T) \mathbf{U}$. Assume that $B_{ii} \geq 0$ for any i . Then

$$\text{Tr}((\mathbf{I}_M - \mathbf{Q}\mathbf{Q}^T) \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T) \geq \left(\inf_{j: B_{jj} > 0} \lambda_j \right) \text{Tr}(\mathbf{B}).$$

Since $\text{Tr}(\mathbf{B}) = \text{Tr}((\mathbf{I}_M - \mathbf{Q}\mathbf{Q}^T) \mathbf{U} \mathbf{U}^T) = \text{Tr}(\mathbf{I}_M - \mathbf{Q}\mathbf{Q}^T)$, we have $\hat{\sigma}^2 \geq (\inf_{j: B_{jj} > 0} \lambda_j)$. Let us prove that $B_{ii} \geq 0$ for any i : for $\mu \in \mathbb{R}^M$, $\mu^T \mathbf{B} \mu = \|\mathbf{U} \mu\|^2 - \|\mathbf{Q}^T (\mathbf{U} \mu)\|^2$ and this term is non negative since $\mathbf{Q}^T (\mathbf{U} \mu)$ is the orthogonal projection of $(\mathbf{U} \mu)$ on the column space of \mathbf{Q} (or equivalently, of \mathbf{S}_M). This equality also shows that

$$\begin{aligned} \{j : B_{jj} > 0\} &= \{j : \exists v \in \mathbf{V}_j, \|v\|^2 > \|\mathbf{Q}^T v\|^2\} \\ &= \{j : \exists v \in \mathbf{V}_j, \|(\mathbf{S}_M^\perp)^T v\|^2 > 0\}. \end{aligned}$$

(iii) Since \mathbf{S}_M is a full rank matrix, \mathbf{R} is invertible. Therefore, from (12), it is trivial that $\hat{\mathbf{K}}$ is positive definite iff $\mathbf{Q}^T (\hat{\Sigma}_M - \hat{\sigma}^2 \mathbf{I}_M) \mathbf{Q}$ is positive definite. Since $\mathbf{Q}^T \mathbf{Q} = \mathbf{I}_r$, we have for any $\mu \in \mathbb{R}^r$, $\mu \neq 0$: $\mu^T (\mathbf{Q}^T \hat{\Sigma}_M \mathbf{Q} - \hat{\sigma}^2 \mathbf{I}_r) \mu > 0$ iff $\mu^T (\mathbf{Q}^T \hat{\Sigma}_M \mathbf{Q}) \mu > \hat{\sigma}^2 \|\mu\|^2$.

Remark.: It can be seen from the proof of (i) that $\hat{\Sigma}_M$ is positive definite iff for any l , \mathbf{W}_l has at least two non null components (say i_l, j_l) such that $D_{i_l} \neq D_{j_l}$.

ACKNOWLEDGMENT

The authors would like to acknowledge Emmanuel De Wailly and Jean-Francois Morlier for their help in data acquisition.

REFERENCES

- [1] 3GPP TSG-RAN WG4 meeting # 62 R4-120908, *Proposal to modeling active antennas*, Katherein, Feb 2012.
- [2] 3GPP TR 36.805 v1.3.0 1, "Study on minimization of drive-tests in next generation networks; (release 9)," 3rd Generation Partnership Project, Tech. Rep., 2009. [Online]. Available: <http://www.3gpp.org/ftp/Specs/html-info/23141.htm>
- [3] J. Mitola, "Cognitive radio: An integrated agent architecture for software defined radio," (*Doctoral dissertation*), Royal Institute of Technology (KTH), Stockholm, Sweden, 2000.
- [4] Y. Zhao, J. Gaedert, K. K. Bae, and J. H. Reed, "Radio environment map enabled situation-aware cognitive radio learning algorithms," *Software Defined Radio Forum (SDRF) technical conference*, 2006.
- [5] A. Alaya-Feki, S. Ben Jemaa, B. Sayrac, P. Houze, and E. Moulines, "Informed spectrum usage in cognitive radio networks: Interference cartography," *IEEE Conference on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pp. 1–5, 2008.
- [6] A. Alaya-Feki, B. Sayrac, S. Ben Jemaa, and E. Moulines, "Interference cartography for hierarchical dynamic spectrum access," *IEEE Conference on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*, pp. 1–5, 2008.
- [7] J. Riihijarvi, P. Mahonen, M. Wellens, and M. Gordziel, "Characterization and modelling of spectrum for dynamic spectrum access with spatial statistics and random fields," *IEEE Conference on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pp. 1–6, 2008.
- [8] 3GPP TR 102 947 V1.1.1, "Reconfigurable Radio Systems (RRS); Use Cases for building and exploitation of Radio Environment Maps (REMs) for intra-operator scenarios," European Telecommunications Standards Institute, Tech. Rep., 2013. [Online]. Available: http://www.etsi.org/deliver/etsi_tr/102900_102999/102947/01.01.01_60/tr_102947v010101p.pdf
- [9] B. Sayrac, J. Riihijarvi, P. Mähönen, S. Ben Jemaa, E. Moulines, and S. Grimoud, "Improving coverage estimation for cellular networks with spatial bayesian prediction based on measurements," *Proceedings of the ACM SIGCOMM workshop on Cellular networks: operations, challenges, and future design*, pp. 43–48, 2012.
- [10] A. Galindo-Serrano, B. Sayrac, S. Ben Jemaa, J. Riihijarvi, and P. Mahonen, "Automated coverage hole detection for cellular networks using radio environment maps," *IEEE Conference on Modeling and Optimization in Mobile, Ad Hoc & Wireless Networks*, pp. 35–40, 2013.
- [11] B. Sayrac, A. Galindo-Serrano, S. B. Jemaa, J. Riihijarvi, and P. Mähönen, "Bayesian spatial interpolation as an emerging cognitive radio application for coverage analysis in cellular networks," *Transactions on Emerging Telecommunications Technologies*, 2013.
- [12] H. Braham, S. Ben Jemaa, B. Sayrac, G. Fort, and E. Moulines, "Low complexity spatial interpolation for cellular coverage analysis," *IEEE Conference on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, 2014.
- [13] —, "Coverage mapping using spatial interpolation with field measurements," *IEEE Conference on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2014.
- [14] N. Cressie and G. Johannesson, "Fixed rank kriging for very large spatial data sets," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 70, no. 1, pp. 209–226, 2008.
- [15] T. S. Rappaport et al., *Wireless communications: principles and practice*. prentice hall PTR New Jersey, 1996, vol. 2.
- [16] C. Phillips, D. Sicker, and D. Grunwald, "Bounding the error of path loss models," *IEEE Conference on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*, pp. 71–82, 2011.
- [17] E. Perahia, D. C. Cox, and S. Ho, "Shadow fading cross correlation between base stations," *IEEE Conference on Vehicular Technology Conference (VTC)*, vol. 1, pp. 313–317, 2001.
- [18] N. A. Cressie, "Statistics for spatial data," 1993.
- [19] J. Riihijarvi and P. Mahonen, "Estimating wireless network properties with spatial statistics and models," *IEEE Conference on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt)*, pp. 331–336, 2012.
- [20] C. Phillips, M. Ton, D. Sicker, and D. Grunwald, "Practical radio environment mapping with geostatistics," *IEEE Conference on Dynamic Spectrum Access Networks (DySPAN)*, pp. 422–433, 2012.
- [21] E. Anderson, C. Phillips, D. Sicker, and D. Grunwald, "Modeling environmental effects on directionality in wireless networks," *Mathematical and Computer Modelling*, vol. 53, no. 11, pp. 2078–2092, 2011.
- [22] R. Ramanathan, "On the performance of ad hoc networks with beam-forming antennas," *Proceedings of the ACM international symposium on Mobile ad hoc networking & computing*, pp. 95–105, 2001.
- [23] Y. Chen, Z. Zhang, and V. K. Dubey, "Effect of antenna directivity on angular power distribution at mobile terminal in urban macrocells: A geometric channel modeling approach," *Wireless Personal Communications*, vol. 43, no. 2, pp. 389–409, 2007.
- [24] E. Anderson, G. Yee, C. Phillips, D. Sicker, and D. Grunwald, "The impact of directional antenna models on simulation accuracy," *IEEE Conference on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks, WiOPT*, pp. 1–7, 2009.
- [25] C. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning*. MIT Press, 2006.
- [26] H. V. Henderson and S. R. Searle, "On deriving the inverse of a sum of matrices," *Siam Review*, vol. 23, no. 1, pp. 53–60, 1981.
- [27] M. Katzfuss and N. Cressie, "Tutorial on Fixed Rank Kriging (FRK) of CO2 data," Department of Statistics Technical Report No. 858, The Ohio State University, Columbus, OH, Tech. Rep., 2011.
- [28] —, "Maximum likelihood Estimation of Covariance Parameters in the Spatial-Random-Effects Model," in *Proceedings of the Joint Statistical Meetings, American Statistical Association*, 2009.
- [29] A. Sengupta and N. Cressie, "Hierarchical statistical modeling of big spatial datasets using the exponential family of distributions," *Spatial Statistics*, vol. 4, pp. 14–44, 2013.
- [30] J. M. Geoffrey and K. Thriyambakam, *The EM algorithm and extensions*. Wiley-Interscience, 2007, vol. 382.
- [31] C. Wu, "On the convergence properties of the EM algorithm," *The Annals of statistics*, vol. 11, no. 1, pp. 95–103, 1983.
- [32] 3GPP TS 136.304 V8.6.0, "Evolved universal terrestrial radio access (E-UTRA); user equipment (UE) procedures in idle mode," 3rd Generation Partnership Project, Tech. Rep., 2009. [Online]. Available: <http://www.3gpp.org/ftp/Specs/html-info/23141.htm>
- [33] S. S. Szyszkowicz, H. Yanikomeroglu, and J. S. Thompson, "On the feasibility of wireless shadowing correlation models," *IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY*, vol. 59, no. 9, 2010.
- [34] "Asset tool," <http://www.aircominternational.com/Products/Planning/asset.aspx>.
- [35] M. Gudmundson, "Correlation model for shadow fading in mobile radio systems," *Electronics letters*, vol. 27, no. 23, pp. 2145–2146, 1991.
- [36] R. Kohavi et al., "A study of cross-validation and bootstrap for accuracy estimation and model selection," *IJCAI*, vol. 14, no. 2, pp. 1137–1145, 1995.
- [37] J. Li and G. Australia, *A review of spatial interpolation methods for environmental scientists*. Geoscience Australia Canberra, 2008, vol. 137.
- [38] F. Gunnarsson, M. N. Johansson, A. Furusk, M. Lundevall, A. Simonsson, C. Tidestav, and M. Blomgren, "Downtilted base station antennas a simulation model proposal and impact on hspa and lte performance," *IEEE Conference on Vehicular Technology (VTC Fall)*, 2008.